ORIENTATION ESTIMATION FOR PLANAR TEXTURED SURFACES BASED ON COMPLEX WAVELETS

N. Anantrasirichai, J. Burn and David R. Bull

Bristol Vision Institute, University of Bristol, Bristol BS8 1UB, UK

ABSTRACT
The gradient of a road or terrain influences the appropriate speed and power of a vehicle traversing it. Therefore, gradient prediction is necessary if autonomous vehicles are to optimise their locomotion. This paper presents a novel texture-based method for estimating the orientation of planar surfaces under the basic assumption of homogeneity. Based on a backprojection technique, we propose a new method for measuring texture consistency in the Dual-Tree Complex Wavelet Transform (DT-CWT) domain. Texture histograms computed with a new anti-aliasing compensation approach are employed to find the rotation of a planar surface. The proposed method performs well with various types of textured surfaces and outperforms other existing methods with significantly reduced computational complexity up to 35%.

Index Terms—orientation estimation, texture, DT-CWT.

1. INTRODUCTION
Numerous scenarios exist where it is necessary or advantageous to estimate surface orientation at a distance from a moving forward-facing camera. Examples include the use of image-based sensors for assessing and predicting terrain geometry in association with the control or navigation of autonomous vehicles. In such cases, the ground ahead of the camera appears as a projective plane in the image. In many real scenarios, the upcoming terrain might not just be flat but may also be oblique and vehicles may need to change speed and gear to ensure safe and clean motion.

Many approaches to extracting 3D information from a scene have been exploited; these include stereo, structure from motion, texture and shading [1]. If only a single image source exists then the methods available are restricted and this is on focus here. This paper presents a novel approach which estimates planar orientation based on texture histograms.

Perspective images of a slanting plane containing a homogeneous surface always exhibit decreasing texture scale as the distance from the observer increases. The density of texture also increases with higher angles to the projection ray [2]. With these properties, texture distortions have been derived in both spatial and frequency representations. Spatial techniques have estimated the orientation using local texels [3] and a Bayes estimator [4, 5]. Backprojection techniques have exploited texture variations of the rectified image [6]. However, these techniques require distinct textures. Some researchers have therefore attempted to estimate shape parameters in the Fourier domain [7–9]. The phase of the sinusoidal component of the surface texture has also been used to derive distortion according geometric projection [10]. Recently, the wavelet transform has been employed to examine the spectral behaviour due to the change of texture density as a function of orientation [11, 12]. Additionally, it has been employed to detect ridges of the textured surface [13]. The wavelet transform overcomes the limitation of the traditional Fourier transform for this application, since it carries information about pixel locations in frequency domain thereby being simple to implement.

Our proposed method estimates the angle of an inclined planar surface in the wavelet domain. Wavelets are an efficient tool for capturing texture information and they have been used in many texture-related applications [14–16]. Here, we employ the Dual-Tree Complex Wavelet Transform (DT-CWT) since this provides near shift-invariance and good directional selectivity [17], which is important for estimating rotation. The DT-CWT has been used in various orientation-related applications [18] and denoising [19]; however, it has never been employed in this specific problem. The novelty of this paper lies in the use of texture histograms to identify the texture consistency of the backprojected image. This exploits the fact that if the surface is correctly rectified, the texture is homogeneous throughout. The minimal distance amongst texture histograms, which are computed from areas throughout the image, should be achieved at the correct rotation of the planar surface. We also introduce an adaptive high-pass frequency gain for anti-aliasing compensation which significantly improve the estimation for real images captured using general commercial digital cameras.

2. WAVELET-BASED PLANAR ORIENTATION ESTIMATION
The Dual-Tree Complex Wavelet Transform (DT-CWT) employs two different real discrete wavelet transforms (DWT) to provide the real and imaginary parts of the CWT. Two fully decimated trees are produced, one for the odd samples and one for the even samples, generated at the first level. This increases directional sensitivity over the DWT and enables discrimination between positive and negative orientations. Six distinct sub-bands exist at each level, the orientations of which are ±15°, ±45°, ±75°, and this is obviously highly beneficial to the estimation of orientation.

Our algorithm starts by backprojecting an image to the surface coordinates, producing a rectified surface image. This rectified image is then transformed to the DT-CWT coefficient space to create a texture map. Subsequently, normalised histograms of areas around the texture map are computed and histogram distances amongst them are recorded. A coarse-to-fine search is performed to find the angle that produces the minimum histogram distance.

2.1. Projective geometry
The projection of the world coordinate $X_w = (X_w, Y_w, Z_w)^T$ of a scene point in the image pixel $(x_i, y_i)$ is $x_i = (x, w, y, w, v)$ with depth $v$. The linear transformation between $X_w$ and $x_i$ is given by $x_i = K[R|t]X_w$, where $K$ is an intrinsic matrix and $[R|t]$
is an extrinsic matrix composed of a 3 × 3 rotation matrix \( R \) and a translation matrix \( t \). Here, we position the camera coordinate at the world coordinate without rotation and translation so that \( x_1 = KX_w \).

A planar surface \( S \) is placed in the coordinate system \( X_s = (X_s, Y_s, Z_s)^T \) on \( x-y \) plane \( (Z_s = 0) \). The origin of \( X_s \) is on the optical axis (\( z \) axis of the camera coordinates) at \( Z_w = Z_o \), i.e. the translation matrix of \( S \) is \( t_s = (0, 0, Z_o)^T \). For orientation of \( S \), most shape-from-texture methods employ the slant-tilt system \((\sigma, \tau)\). The slant is defined as the angle between the surface normal and the \( Z_w \) axis, while the tilt is the angle between the \(-x_1\) axis and the projection of the \( n_s \) onto the image plane \( [2, 4, 21] \). Instead, we employ the rotation simply represented by the rotations about the \( x \) and \( y \) axes. This is computed using two basis rotations as in Eq. 1.

\[
R_{sw} = \begin{bmatrix}
\cos \theta_y & 0 & -\sin \theta_y \\
0 & 1 & 0 \\
\sin \theta_y & 0 & \cos \theta_y
\end{bmatrix} \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \theta_x & -\sin \theta_x \\
0 & \sin \theta_x & \cos \theta_x
\end{bmatrix}
\]

We have chosen this system because it is straightforward for converting each image pixel \( x_1 \) to \( X_s \) on \( S \) using the relationship as \( x_1 = K R_{sw} X_s + K t_s \). \( S \) can be seen as the rectified image of the planar terrain image. Also, the estimated surface \( \hat{S} \) can be constructed by backprojecting \( x_1 \) to \( X_s \) with defined \( \theta_x \) and \( \theta_y \).

In a real scene, a terrain surface is generally large, i.e. the scale of \( X_s \) is significantly larger than that of \( x_1 \) and this can lead to high computation time. Therefore, \( S \) is projected onto the virtual image plane \( x_{a4} \) which is parallel to the \( x-y \) plane of the surface coordinate. That is, \( x_{a4} = f/X_s \), where \( f \) is focal length (which generally can be obtained from EXIF header data in the image file).

### 2.2. Texture map

The texture map \( M_{\theta_x, \theta_y} \), created from the reprojectation of the surface image \( \hat{S} \) using \( \theta_x \) and \( \theta_y \), is defined from the highpass coefficients of the DT-CWT. The DT-CWT decomposes \( \hat{S} \) (size \( w \times h \)) into \( l = 1, \ldots, L \) scales, where each scale comprises 6 subbands \( \Psi_l^i, t = 1, \ldots, 6 \) and each subband is of size \( \frac{w}{2^t} \times \frac{h}{2^t} \). We denote \( U_l(\bullet) \) as the 2D Gaussian kernel interpolation with up-sampling by a factor of \( 2^t \). We also apply a mask \( B \) to exclude the large highpass magnitudes which occur around the boundary of the reprojected surface \( \hat{S} \). The erosion operation with a disk size of \( 2^t \) is performed. The \( M_{\theta_x, \theta_y} \) is then defined as Eq. 2.

\[
M_{\theta_x, \theta_y} = B \cdot \sum_{l=1}^{L} U_l(\sum_{t \in T} |\Psi_l^i|)
\]

where \( T \) is the set of subbands employed to create the texture map. This paper aims to estimate the orientation of terrain for locomotive applications, i.e. \( \theta_x \) is always positive and \( \theta_y \) is likely to be between \(-45^\circ \) and \(45^\circ \). Hence, we employ only 4 subbands, ignoring the horizontal direction, i.e. \( T \in \{\pm 45^\circ, \pm 75^\circ\} \). All results in this paper were generated using \( L=3 \). Example is shown in Fig. 1.

**Anti-aliasing compensation:** Slanting textured images captured using digital cameras may suffer from blur at a distance, especially where high frequencies dominate. This is because most manufacturers insert an optical lowpass filter between the lens and imaging sensor to prevent moiré patterns. The filter slightly blurs the fine detail that is similar in frequency to the resolution of the sensor and creates a tradeoff between sharpness and aliasing. In particular, when projecting the surface onto the rectified view, distant areas exhibit loss of high frequencies. Deconvolution methods may be employed to restore those details [16], but system complexity could be significantly increased. We therefore propose a simple technique using an adaptive distance-related gain.

The 1D projective frequency \( F_p(x) \) at instant position \( x \) on the image plane capturing the line with slant \( \theta_x \) is derived in [13] as shown in Eq. 3, where \( f_0 \) is the frequency on surface \( S \).

\[
F_p(x) = \frac{f_0}{\cos(\theta_x)(f - x \tan(\theta_x))^2}
\]

This implies that more pixels are affected by the anti-aliasing filter at \( x > 0 \) than those at \( x = 0 \) by \( F_p(x)/F_p(0) \). We therefore exploit this concept to boost the magnitudes of high frequency coefficients with a gain \( G_i(x) \) as shown in Eq. 4. \( \alpha \) is a gain control that must be less than 1 because the distribution of the filter is not uniform. Empirically, \( \alpha = \frac{1}{4} \) performs well for most natural terrain images. The term \( 2^{(L-1)} \) is also included so that noise at fine scales is not enhanced by the gain term.

\[
G_i(x) = 1 + \frac{1}{2^{(L-1)}} \max(0, \left[ \frac{f}{f - x \tan(\theta_x)} \right]^{-2\alpha} - 1)
\]
The rectified image is the reconstructed planar surface located on the \(x-y\) plane of the coordinate system \(X_s\) as described in Section 2.1. When we view this \(x-y\) plane from the top (\(z\) axis), the texture surface must reveal texture consistency under the assumption of homogeneity. This section describes a histogram-based approach used for measuring texture consistency on the estimated \(\hat{S}\).

**Texture histograms:** Texture histograms are constructed from 9 regions of the texture map \(M_{\theta_x,\theta_y}\). These regions are arranged as a \(3 \times 3\) grid on \(M_{\theta_x,\theta_y}\) as shown in Fig. 1 (a5) and (b2). Each side of the quadrilateral is divided into 6 equal lengths resulting in 5 points. The intersections of lines 1, 3 and 5 joining between the opposite sides are used as the centres of the 9 regions. The region size \((r_w \times r_h)\) is defined adaptively according to the size of \(S\) as follows:

\[
(r_w, r_h) = \left( \frac{\sum B}{TR_B}, \frac{\sum B}{TC_B} \right)
\]

where \(C_B\) and \(R_B\) are the numbers of columns and rows of mask \(B\). The positions of some regions may be shifted slightly such that the whole region is located inside \(B\). Subsequently, the normalised histogram \(h_r\) of region \(r\) \((r = 1, \ldots, 9)\) is computed using pre-estimated bin size. We approximate bins of the histogram from the texture map of the rectified image reconstructed using \(\theta_x = 45^\circ\) and \(\theta_y = 0^\circ\), i.e. \(M_{45^\circ,0^\circ}\). If \(K\) is the total number of bins, the interval for each bin \(k\) is defined as \(v_k \in \left( \frac{(k-1)M}{K=1}, \frac{M}{K=1} \right), \quad k = 1, \ldots, K\), where \(M = \max(M_{45^\circ,0^\circ}) + 0.01\).

**Cost function:** Texture histograms such as those shown in Fig. 2 left and middle are computed from the correct and incorrect rectified images shown in Fig. 1. All histogram shapes for the correct rectified image are more similar than those for the incorrect rectified image. Histogram distance is employed to measure the texture consistency amongst regions of the rectified image. A small distance indicates high correlation between regions, so a small summation of histogram distances amongst the 9 regions indicates texture consistency throughout \(\hat{S}\). The cost function \(D\) could be calculated from all pairwise combinations of histograms, but this consumes excessive computational resource. We therefore calculate the average histogram \(\bar{h}\) and find the histogram distance \(d_r\) between \(h_r\) of region \(r\) and \(\bar{h}\). The final cost function \(D\) is defined as Eq. (7):

\[
D = \sum_{r=1}^{9} d_r
\]

We have tested several histogram distance metrics in order to find the best fit to our problem. The bin-to-bin distances include Euclidean distance, City block metric and Chi-square distance. For, cross-bin distances, we employ Earth Mover’s Distance (EMD) implemented in [22]. The results of these are presented in Section 3.

**2.4. Initial orientation**

We estimate the start point of the search process to find the least cost \(D\) using an inverse of \(F_p(x)\) (Eq. 3), \(T_p(x)\), which is a parabolic function of \(x\). That is,

\[
T_p(x) = \frac{\cos(\theta_x)(f - x \tan(\theta_x))^2}{fZ_0f_0} = ax^2 + bx + c
\]

\(T_p(x)\) is then normalised, \(\hat{T}_p(x)\), to exclude the unknown parameter \(f_0\). The position \(x\) is also scaled to \(\hat{x}\) where \(\hat{x} \in [0, 1]\) so that only the coefficient of \(\hat{x}^2\) is important – the quadratic coefficient indicates the shape of the parabola. Plot of \(\hat{T}_p(\hat{x})\) is shown in Fig. 2 (right) compared to the real \(T_p\) constructed from the synthetic textural images at various angles. These images are generated from defined sinusoids, so the actual local frequencies are known. We construct the structure map \(M_{struc}\), described in Eq. 9, to generate the 1D signals, \(T_p(x)\) and \(T_p(y)\).

\[
M_{struc} = U_3 \left( \sum_{l \in T} |\Psi_1^l| \right)
\]

The highpass subband at level 3 is employed since this reveals clear structure information without noise interference. \(T_p(x)\) and \(T_p(y)\) are an inverse of a column-wise sum and an inverse of a row-wise sum of \(M_{struc}\), respectively. After normalising \(T_p(x), x \in [-\frac{1}{2}, \frac{1}{2}]\) and \(T_p(y), y \in [-\frac{1}{2}, \frac{1}{2}]\), a polynomial curve fitting with degree of 2 is independently applied to find coefficients \(\{a_x, b_x, c_x\}, \varphi \in \{x, y\}, \kappa \in \{w, h\}\) of \(\hat{T}_p(\hat{x})\). The initial \(\theta_x\) is obtained from Eq. 10.

\[
\theta_x = \tan^{-1}\left(\frac{2\alpha f_1}{\kappa}\right)
\]
3. RESULTS AND DISCUSSION

The proposed method was tested using both synthetic and real images. For real images, two types were used. The first was generated from the top-view images and projected to different angles so that the groundtruths were known, while the second type was acquired from real scenes. Performance comparisons were made with the methods of the homogeneity revisited (HR) [6], Galasso & Lasenby (GL) [8] and Hwang, Lu & Chung (HLC) [13].

Synthetic images: To construct synthetic images, the rectified images with homogeneous texture were generated from checkerboards of size 1000×1000 pixels. The block sizes of the checkerboard where set to 5×5, 20×20, 35×35 and 50×50 pixels. These are also rotated with angles of 0°, 25°, 45° and 65°. Gaussian low-pass filters with standard deviations of σ=0.5 and 5 were then applied. This created a total of 32 images (such as that shown in Fig. 1 (a1)). Subsequently, with defined values of θx and θy, the slant surface image was constructed assuming a focal length of 150 and Z0=200. Table 1 shows the average results for each of the 32 synthetic images. It can be observed that the Euclidean distance metric gives the best results. It also has the lowest computational cost of all metrics. According to the accuracy and computational time, the Euclidean metric was used in the test with real images.

Real materials: Textured images from the Brodatz database [23] were used to synthesise the oblique surfaces. Rotations were applied to the same images as in [8], with which our results were compared. These images can be seen in [8] (Fig. 3 (a)-(g)). As the Brodatz images are represented in a different projective geometry, slant σ and tilt τ are converted to our system using θx = \cos^{-1}\left(\frac{\tan^2\tau \cos^2\sigma}{1 + \tan^2\tau}\right) and θy = \tan^{-1}\left(\frac{\tan(\tau)}{\cos\sigma \cos\theta_y}\right). The results are shown in Table 2 (we set the search precision to 0.05°). The proposed method based on Euclidean distance gives the best estimation of textured planar surfaces. Based on a backprojection technique, the texture consistency of the rectified image using texture histograms. The proposed method is used with real scenes, the achieved average error of less than 4° is sufficient for prediction of the terrain gradient ahead of an autonomous robot or vehicle.

4. CONCLUSIONS

This paper has presented a new method for estimating the orientation of textured planar surfaces. Based on a backprojection technique, we have proposed a new wavelet-based method to measure the texture consistency of the rectified image using texture histograms. An anti-aliasing compensation is also employed. Results show that the proposed method achieves high accuracy with an average error of less than 1° for both synthesised images and real textures from the Brodatz database. Our approach outperforms the existing methods, while reducing computation by 35%. Although the accuracy decreases when the proposed method is used with real scenes, the achieved average error of less than 4° is sufficient for prediction of the terrain gradient ahead of an autonomous robot or vehicle.

### Table 1. Orientation estimation for synthetic images

<table>
<thead>
<tr>
<th>Method</th>
<th>θx</th>
<th>θy</th>
<th>θx</th>
<th>θy</th>
<th>θx</th>
<th>θy</th>
<th>mean error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean</td>
<td>15.0°</td>
<td>-3.0°</td>
<td>45.0°</td>
<td>8.0°</td>
<td>73.0°</td>
<td>-5.0°</td>
<td>(\hat{\theta}_x, \hat{\theta}_y)</td>
</tr>
<tr>
<td>Cityblock</td>
<td>14.42°</td>
<td>-2.07°</td>
<td>43.52°</td>
<td>7.51°</td>
<td>72.32°</td>
<td>-4.96°</td>
<td>0.72 0.38</td>
</tr>
<tr>
<td>Chi-square</td>
<td>14.43°</td>
<td>-2.20°</td>
<td>43.48°</td>
<td>8.10°</td>
<td>73.04°</td>
<td>-5.01°</td>
<td>0.54 0.35</td>
</tr>
<tr>
<td>EMD</td>
<td>14.52°</td>
<td>-2.52°</td>
<td>43.44°</td>
<td>7.49°</td>
<td>73.05°</td>
<td>-4.95°</td>
<td>0.52 0.38</td>
</tr>
<tr>
<td>HR [6]</td>
<td>13.88°</td>
<td>-3.76°</td>
<td>47.20°</td>
<td>7.54°</td>
<td>74.82°</td>
<td>-6.98°</td>
<td>1.85 1.20</td>
</tr>
<tr>
<td>GL [8]</td>
<td>14.44°</td>
<td>-2.00°</td>
<td>42.84°</td>
<td>9.75°</td>
<td>73.33°</td>
<td>-4.78°</td>
<td>1.05 0.78</td>
</tr>
</tbody>
</table>

### Table 2. Orientation estimation errors for real materials

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}_x), (\hat{\theta}_y)</td>
<td>(\hat{\theta}_x), (\hat{\theta}_y)</td>
<td>(\hat{\theta}_x), (\hat{\theta}_y)</td>
<td>(\hat{\theta}_x), (\hat{\theta}_y)</td>
<td></td>
</tr>
<tr>
<td>Cityblock</td>
<td>D20</td>
<td>D52</td>
<td>D25</td>
<td>D57</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.67</td>
<td>0.07</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>0.53</td>
<td>1.06</td>
<td>0.75</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>0.75</td>
<td>0.20</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>1.20</td>
<td>0.61</td>
<td>0.23</td>
</tr>
<tr>
<td>Grass</td>
<td>D65</td>
<td>D82</td>
<td>D84</td>
<td>D95</td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>0.83</td>
<td>0.28</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>1.24</td>
<td>0.25</td>
<td>0.13</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.80</td>
<td>1.55</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>3.48</td>
<td>1.96</td>
<td>2.38</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td>1.22</td>
<td>0.84</td>
<td>0.58</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.43</td>
<td>0.07</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>1.39</td>
<td>2.46</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>0.38</td>
<td>0.80</td>
<td>4.10</td>
<td>1.98</td>
</tr>
</tbody>
</table>

### Table 3. Orientation estimation errors for real scenes

<table>
<thead>
<tr>
<th>Image</th>
<th>without (G_l)</th>
<th>with (G_l)</th>
<th>GL [8]</th>
<th>HLC [13]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bricks</td>
<td>5.25</td>
<td>3.75</td>
<td>3.00</td>
<td>3.75</td>
</tr>
<tr>
<td>Tarmac</td>
<td>3.00</td>
<td>3.25</td>
<td>2.25</td>
<td>4.50</td>
</tr>
<tr>
<td>Grass</td>
<td>7.25</td>
<td>5.25</td>
<td>4.00</td>
<td>6.50</td>
</tr>
<tr>
<td>Sand</td>
<td>6.50</td>
<td>4.75</td>
<td>3.00</td>
<td>5.25</td>
</tr>
</tbody>
</table>

Fig. 3. Rectified images of glass and bricks for (a) near areas and (b) far areas, constructed (a1) and (b1) without gain \(G_l\), (a2) and (b2) with gain \(G_l\).
5. REFERENCES


